引用格式：徐芳，李晓轩. 国际科技评价改革十年评述. 中国科学院院刊, 2024, 39(1): 121-130, doi: 10.16418/j.issn.1000-3045.20231217002.
Xu F, Li X X. Review on reform of international research evaluation in past decade. Bulletin of Chinese Academy of Sciences, 2024, 39(1): 121-130, doi: 10.16418/j.issn.1000-3045.20231217002. (in Chinese)

国际科技评价改革十年评述

徐 芳 李晓轩*

1 中国科学院科技战略咨询研究院 北京 100190
2 中国科学院大学 公共政策与管理学院 北京 100049
3 中国科学学与科技政策研究会 科技管理与评价专委会 北京 100080

摘要 2013年5月，以《科研评价的旧金山宣言》（DORA）发布为标志，拉开了国际科技评价改革的序幕。10年来，国际科技评价改革经过大量理论准备和倡议，正在走向实践，40多个国家的350多家机构签署《改革科研评价的协定》并进行改革试点，效果初显。文章从改革目标、举措及案例3个方面，对国际科技评价10年改革进行了系统总结和分析，并从与我国科技评价改革比较的角度，得出研究结论和启示。

关键词 国际，科技评价，改革，旧金山宣言

DOI 10.16418/j.issn.1000-3045.20231217002

CSTR 32128.14.CASbulletin.20231217002

2013年5月，《科研评价的旧金山宣言》（以下简称DORA或《旧金山宣言》）正式发布[1]，旨在解决逐渐兴起的“以刊评文，以文评人”问题。《旧金山宣言》得到国际科学共同体广泛认同与反响；以此为旗帜，许多国际学术组织、学术年会、高校与研究机构开始讨论科技评价改革。同时，成立了DORA科技评价联盟、科技管理国际联盟（INORMS）科技评价工作组等新的国际组织致力于推动科技评价改革。10年来，国际科技评价改革不断深入，逐步从理念层面的提倡和讨论落实到众多科研机构的实践探索之中，效果初显。

笔者曾于2022年发表过《科技评价改革十年评述》一文[2]，对我国科技评价10年改革进行过总结。认为，以“三评”改革和破“四唯”为代表的我国科技评价改革正处在关键时刻。虽然目前在清理“四唯”上已取得初步效果，以论文等指标进行简单量化评价的现象明显好转。但是，“立新标”还在半路上，特别是科技评价改革所要引导的对卓越的价值追求还

*通信作者
资助项目：国家自然科学基金面上项目（72174191），中国科学院战略研究与决策支持系统建设专项（GHJ-ZLZX-2023-22）
修改稿收到日期：2024年1月7日

远未形成。对此，如何谋划下一步科技评价改革的目标和举措是迫切需要回答的重要课题。笔者作为国际科技评价组织的任职者①和经历者，通过对国际科技评价改革10年进行系统的梳理、分析和比较，得出相应结论和启示，以期起到他山之石的效果。

为避免引起歧义，对文中2个概念进行说明：① 本文中提及的国内、国际的科技评价均指针对高等学校、科研机构（含科研资助机构）开展的以基础研究为主的科研评价，包括论文、人才、项目、机构等的评审评价，尽管英文一般被称为“research assessment”或“research evaluation”，但是，为了与国内语境保持一致，本文沿用“科技评价”而不是“科研评价”。② 国际科技评价改革主要是指欧洲和北美等地区的传统科技强国主导、目前在国际上形成较大影响的科技评价改革，包括改革目标、改革的组织推动过程、改革的基本理念及开展的改革实践等。

1 国际科技评价改革要解决什么问题?

一般而言，欧洲和北美等传统科技强国因为其深厚的科学文化底蕴，科学价值标准与追求传统上是不错的。但是，随着科技自身的发展及其地位的提升，其科技评价也面临新的问题与挑战，必须与时俱进。归纳起来，国际科技评价改革要应对的问题或者说要达成的目标大致有3个方面。

（1）避免文献计量学方法在科技评价中的不当使用。在美国学者Garfield[3]提出利用参考文献追踪科学进展的引文分析方法后，美国科学信息研究所（ISI）于1963年开始通过引文分析筛选期刊，形成科学引文索引（SCI）数据库，从而为在科研评价中应用文献计量学方法提供了基础。文献计量学方法的引入，一方面为科技评价提供了证据支撑；另一方面助推了“以刊评文”的逐步兴起——“发表在哪儿比发表什么更重要”，这无疑对科研产出的质量、完整性和多样性产生不利影响。如何避免文献计量学方法的不当使用，成为国际科学共同体面临的重要挑战。

（2）重视科学对经济社会的影响力（impact）评价。随着科技在国家经济社会发展、国家安全等方面地位的提升，科技竞争愈演愈烈，世界各国一方面加大科技投入，另一方面也更加关注科技投入对本国创新发展的效率和效果。原来科学投入“只问耕耘，不问收获”的线型模型被质疑，科学对经济社会的影响力评价逐渐成为科技评价的核心内容。影响力评价的引入带来2个方面的挑战：① 科学共同体难以形成共识，很多科研人员不认可影响力评价，认为这种边界模糊、容易自我吹嘘的评价会助长学术不端、损害学术质量。② 准确评价影响力难度太大，难以找到科学的指标、数据来源与评价方法。这2个问题也是国际科技评价界讨论的热点问题。

（3）适应开放科学、基于人工智能的科学研究等新范式发展。以数据共享为基础的开放科学（open science）[4]近年来在欧美盛行并逐渐影响全球。开放科学运动兴起的同时要求改革科技评价系统，以提高公开性和透明度。但是，如何将传统上以个人创造活动为主的科学研究，转变为数据共享的、体现大规模协作的集体性科研活动并非易事，需要各方面共同努力。例如，联合国教育、科学及文化组织（UNESCO）提出了关于开放科学的建议[5]，其中包括为其成员开发的“开放科学工具包”，帮助他们审查和改革科研职业生涯的评估标准。人工智能的迅猛发展也将对科技评价产生深远影响，“人工智能驱动的科学研究”（AI for Science，简称AI4S）成为新的科学范式[6]。各国都致力于抢占这一范式制高点，也需要通过科技评价予以激励和引导。同时，AI4S在促进

① 笔者担任DORA科技评价联盟执委和科技管理国际联盟科技评价工作组成员。

科技发展、减轻科学家负担的同时，可能强化数据预测技术而带来风险与偏见，也对改革科技评价提出了新挑战。不过，在这方面目前虽然有较多提及，但目前还没有成为10年来国际科技评价改革的重点。

就以上3个方面而言，这次国际科技评价改革要解决的紧迫问题或者说核心目标是前2个方面，即文献计量学方法不当使用，以及影响力评价。这与我国是相似的，第1个问题即文献计量学方法不当使用方面的挑战在我国尤其严峻。这是因为，相比于传统科技强国，由于科学文化薄弱、人情因素过多等原因，我国的同行评议系统还不够健全，这导致科技评价中对论文发表的刊物的影响因子、论文自身的引用量及论文数量等量化指标更加倚重。第2个问题即如何推动影响力评价跟我国正在推动的科技成果“五元价值”评价是相似的。但是，我国破“四唯”②中的奖项、学历、职称、人才“帽子”等，更多是我国特色，对于一些传统科学强国而言这些基础性问题并不显著。

2 国际科技评价改革采取了什么举措?

2.1 国际科技评价改革的若干重要举措

从路径来看，国际科技评价改革由科学共同体主导，主要采用自下而上方式开展。国际科技评价改革启动的标志性事件是2013年5月《旧金山宣言》的发布。该宣言的初稿是相关学者和编辑在2012年末旧金山召开的美国细胞生物学学会（ASCB）年会期间，针对期刊影响因子在科技评价中不当使用的弊端所提出的。《旧金山宣言》发布后，许多国际学术组织、学术年会、高校与研究机构纷纷跟进，并成立了DORA科技评价联盟等新的国际组织共同致力于推动科技评价改革。2023年5月，包括中国在内的全球许多国家分别举行了《旧金山宣言》发布10周年纪念活动③。

10年来，国际科学共同体在推动科技评价改革上做了大量形式多样的工作，包括发布宣言、倡议、声明；组织学术年会交流、专题研讨、项目研究；形成研究专报、科技评价方法框架、好的评价案例、科技评价试点协定等。本文梳理了其中14项比较重大的举措（表1）。

2.2 国际科技评价改革的主要效果

（1）在全球范围内形成科技评价改革共识。截至2024年1月4日，已有3 078个组织和21 339名个人签署了《旧金山宣言》，其中包括来自中国的15家机构。2022年，“推进科研评价联盟”（CoARA）正式成立，并发布了《改革科研评价的协定》，来自40多个国家的350多个组织签署了协定。科技评价改革日益在全球范围内形成共识。

（2）经过科学共同体各方力量共同努力，科技评价改革的“图谱”逐渐清晰化。比如，《旧金山宣言》提出破除“以刊评文”；《莱顿宣言》进一步提出要纠偏“量化评价”；《量化指标潮流》报告进一步明确量化评价的作用及规范；SCOPE框架定义负责任评价的过程等。不同的学术组织针对不同问题提出科技评革的不同方面，拼凑成一个比较完整的“图谱”。最后，这张改革“图谱”被冠以“负责任科研的评价”（responsible research assessment）的标识，逐步成为科技界的共同用语。

（3）科技评价改革正在从理念层面走向实践。目前，签署《旧金山宣言》的3 000个多组织正在或者已经落实避免“以刊评文”要求。签署《改革科研评价的协定》的300多个组织（包括资助机构、高校和科研机构）正在进行科技评价改革试点，并且经常组

② 以后破“四唯”增加了破“唯帽子”，发展为破“五唯”，即破：唯论文、唯奖项、唯学历、唯职称、唯帽子。

③ 中国区活动由中国科学学与科技政策研究会科技管理与评价专委会于2023年5月19日举办。

表1 2013—2022年国际科学共同体推动科技评价改革的14项重大举措[1]

Table 1 14 major measures to promote research assessment reform from 2013 to 2022 by international research community

序号	年份	举措	发起者	核心内容
1	2013年	《科研评价的旧金山宣言》(DORA)	DORA科技评价联盟	针对期刊影响因子的使用，提出了18条建议，包括总体建议和针对资助机构、研究机构、出版机构等不同主体的建议
2	2015年	《莱顿宣言》(*Leiden Menifeso*)[7]	莱顿大学科学技术研究中心(CWTS)	提出了规范使用定量指标的十大原则
3	2015年	《量化指标潮流》报告(*Metric Tide*)[8]	英格兰高等教育基金委员会(HEFCE)	对定量指标在英国卓越研究框架(REF)中的使用进行评估与反思，提出建议
4	2016年	《人文科学定量指标的15个关键原则》(*15 Key HuMetricsHSS Principles*)[9]	HuMetricsHSS工作组	提出了基于价值观的评估框架，特别针对人文学科和社会科学，提出了有益的评估指标，以指导学术评估，促进对学术界价值观制定方法的培养
5	2017年	《下一代指标：负责任的开放科学指标和评估》(*Next-generation metrics: responsible metrics and evaluation for open science*)[10]	欧盟委员会	提出未来的科研评估模型应将专家评议和定量指标相结合，秉承透明、严谨的原则，利用好现有指标，衡量"真正重要"的内容，使评估指标具备开放性、互操作性
6	2017年	《转型中的科学》报告(*Science in Transition*)[11]	欧盟委员会	强调了科技评估模型指标融入开放科学实践中的重要性
7	2018年	负责任评估的SCOPE框架[12]	科技管理国际联盟科技评价工作组(INORMS-REW)	提出了负责任评估的5个步骤：评估价值观(start)、情景考虑(context)、评估选项(options)、深入调研(probe)、元评估(evaluation)
8	2018年	《向开放科学过渡的研究评估路线图》(*Roadmap on Research Assessment in the Transition to Open Science*)[13]	欧洲大学联盟(EUA)	支持成员机构制定涵盖质量、潜力和影响的负责任评价方法
9	2018年	《科学卓越、出版模式、评估和开放科学》(*Scientific Excellence, Publishing models, assessment and open science*)[14]	全球青年科学院(GYAWG)卓越科学工作组	处于职业生涯早期的年轻科研人员共同提出了关于改进科研评估过程的15项建议
10	2018年	《同行/价值评议原则声明》(*Statement of Principles on Peer/Merit Review*)[15]	全球研究理事会(GRC)	提出了负责任专家评审的原则
11	2019年	《科研人员评价的香港原则》(*Hong Kong Principles for Assessing Researchers*)[16]	第六届科研诚信国际会议	提出将学术诚信作为科研人员考核评价的重要标准
12	2020年	《科学欧洲关于研究评估过程的立场声明》(*Science Europe Position Statement on Research Assessment Processes*)[17]	科学欧洲(Science Europe)	基于《旧金山宣言》和《莱顿宣言》提出了若干补充性建议
13	2020年	《负责任科研评估会议总结报告》(*RRA Conference Summary Report*)[18]	全球研究理事会(GRC)	在全球研究资助机构内形成负责任评估的共识
14	2022年	《改革科研评价的协定》(*The Agreement on Reforming Research Assessment*)[19]	推进科研评价联盟(CoARA)	提出承认科学工作的多样性、承诺科研评价的合理性、抛弃"论文成功学"、避免评价偏见等4项核心承诺及6项支持性承诺

注：表1整理过程中参考了文献[20]

Noted: Table 1 was compiled with reference to Ref. [20]

织各种形式的试点经验交流。

（4）形成了关于科技评价的一些基本判断。包括对科技评价正、反两方面作用，定量评价与定性评价之间的关系，启动评价的前提条件，以及提升评价数据质量的技术等。这些理性认识对我国有重要启示意义。

3 国际上的“三评”改革实践案例

如前文所述，国际科技评价改革正在从理念层面走向实践，以下进行实践案例分析。鉴于我国当前科技评价源于国家“三评”改革文件，在此也分别选择人才评价、项目评审和机构评估3个方面的案例进行分析。

3.1 比利时根特大学人才评价改革

比利时根特大学（Ghent University）较早注意到基于文献计量学方法的定量评价对研究文化带来的系统性损害，认为定量评价助长了“发表在哪儿比发表什么更重要”的文化。2013年发布后，根特大学签署了《旧金山宣言》。之后，又签署了《改革科研评价的协定》。随之着手对科教人员晋升评价等人才评价进行改革，以维护根特大学一贯倡导的多元化文化，同时消除教师对量化评价越来越大的不满，力图创建一个强调追求卓越研究的共同价值观、富有挑战性、高质量且具激励性的职业框架。

在学校管理层和科教人员的共同努力下，2016年11月，根特大学发布了《根特大学评价研究愿景声明》[21]，提出科研评价必须遵守8项原则。2017年进一步公布了科研评价中定量指标的使用指南[22]。根据这2项政策，2018年根特大学建立了一种全新的教师评价和晋升模式，重新将“责任”和学术自由还给教授级教职员。根据新的评价体系，根特大学对教师的评价不再只看科研产出，而会从更加定性、综合且以人为本的视角进行评价。评价以5年为周期，包括初期的证据性评价、中期反馈访谈和末期的访谈式评价。评价内容包括在研究、教学、社会参与度、管理和领导力方面最重要成就的叙述性展示（而非使用可衡量的定量标准），以及未来5年的意向规划。

3.2 美国国立卫生研究院（NIH）项目评审改革

美国国立卫生研究院（NIH）是美国最主要的医学研究和资助机构，每年都会资助大量新项目以推动相关领域的发展。NIH在签署了《旧金山宣言》后，着手进行项目评审改革，以消除已有评审中存在的量化问题与偏见，同时适应开放科学的发展。改革主要包括3个方面。

（1）修改评审规则[23]。新规则要求对于研究者和研究环境的评估必须放在研究项目的背景下来考虑而不是像之前被单独评分。无论是人员还是机构，评审标准不再是“越强越好”而是“能胜任即可”；如果评审专家认为人员或机构能力不足，则需要给出具体说明。新标准的“够用”原则试图尽量解决声望偏见，将关注点更多地放在研究课题本身上面而非机构声誉。

（2）修改项目申请所用的简历格式或“专家介绍”[24]。在专家介绍中添加小段篇幅，由申请人简要描述其最重要的科学成就，以转移项目评审专家对以往研究论文发表期刊的关注度。

（3）出台数据管理和共享新政策。自2023年1月起，要求每年受NIH资助的30万名研究人员和2 500个机构中的大多数在其拨款申请中阐明数据管理和共享（DMS）计划[25]。DMS计划中应包括分析数据所需的软件或工具的细节、何时何地公布原始数据，以及访问或分发该数据的任何特殊考虑，并对数据共享的任何限制或例外情况说明理由，以促进开放科学发展。

3.3 英国大学评价改革

2014年，英国对原有大学科研评估考核（RAE）体系进行较大幅度改革，形成新的科研卓越框架（REF）。相较于以往的评估体系RAE，REF最大的改革亮点在于：① 引入了文献计量学评价指标，为同行评议提供参考；② 探索了影响力评价方法[26]，以展示英国大学研究对社会的真正影响，强调科学研究给现实世界带来的利益。由于影响力评价难度较大，英国为此进行专门研究，开发了针对不同类型学科科研成果的影响力指标。

2014实施的REF（REF 2014），在实现以评估结果对大学进行资源配置的同时，也不可避免地将竞争压力通过大学传导至基层学术组织，特别是定量指标的引入加剧了对教师个人科研行为的影响。2015年，受英国高等教育基金委员会委托，以James Wilsdon教授为首的研究组对定量指标在REF中的作用进行了独立评估。为此，研究组启动了负责任评价指标专题论坛，重点关注与讨论在科技评价中如何用好定量指标的问题。最后，研究组发布了题为《量化指标潮流》的研究报告，对使用定量指标给出了肯定判断并提出改进建议。

2022年，英国REF的领导机构——英格兰研究院（Research England）签署了《改革科研评价的协定》，要对REF继续进行改革，旨在整个高校科研生态系统重新建立起一种负责任、包容、多样的科研文化。由此，英国启动了“未来科研评价计划”[27]，旨在对刚实施完毕不久的REF 2021未来变革方案进行更深入研究。根据2023年6月发布的新一轮顶层设计方案，REF 2028的政策重心将从“科研绩效激励”调整为“科研文化建设”，并且全面重塑科研环境、科研成果和科研影响3个评价维度，以提升高校对建设健康科研文化的重视程度。

3.4 国际上“三评”改革实践的主要经验

*（1）理论与实践结合。*试点机构充分利用科技评价10年改革期间国际科学共同体构建的一系列理论和方法体系指导实践，理论与实践紧密结合。

*（2）保持与其他试点机构的交流共享。*试点机构一般会签署《旧金山宣言》和《改革科研评价的协定》，置身于试点集体之中，与科技评价研究的学者以及其他试点机构保持互动交流。

*（3）在评价方法上保持不断创新和完善，而不奢望一蹴而就。*例如，REF自2008年提出引入科研影响力评价，期间花费了大量的人力和时间研究开发影响力的内涵、评价标准、评价方法、专家手册等，直至2014年才应用于实际评价之中，而且现在依然在研究和完善之中。

*（4）试点机构体现了充分的改革自主性。*各机构响应国际科学共同体的倡议启动改革，是出于自身在理念上的认同和实际需求，完全是自主的，而不是源自政府行政要求。

4 结论与启示

4.1 结论

国际科技评价10年改革值得总结的内容很多，本文主要从与我国科技评价改革比较角度，得出3个方面结论。

*（1）国际科技评价改革目标与我国相似。*本次国际科技评价改革的核心目标有2个：① 破“以刊评文”，这与我国破“四唯”之中破“唯论文”的导向是一致的；② 立对经济和社会贡献的“影响力”评价，这与我国强调科技成果的五元价值[28]是一致的。但是，我国破“四唯”中破奖项、学历、职称、人才“帽子”等，主要是我国特色。对于国际上的传统科学强国而言，科技奖项、人才“帽子”等基础性评价问题并不明显。

*（2）国际科技评价改革路径与我国差异较大。*国际科技评价改革主要由科学共同体主导，采用自下而上的方法，通过科学共同体发布宣言、倡议、评价方法体系、签署承诺协议、实践案例总结与分享等方式推动，政府很少直接介入。相反，我国更多采用自上而下方法，政府在科技评价改革中发挥主导作用，通过发布改革的政策文件和要求推动改革，科学共同体的作用发挥有限。

*（3）国际科技评价10年改革经验值得借鉴。*与自上而下的我国科技评价改革执行力强的特点不同，国际科技评价改革更加关注对科技评价规律的认识。由此，国际科技评价改革达成了比较系统的共识，值得借鉴。当然，国际科技评价改革所形成的许多共识跟

我国在实践中秉持的理念是相似的，有些甚至有异曲同工之妙。国际科技评价改革的共识包括：《旧金山宣言》提出的关注论文本身而不是期刊的倡议及对不同主体做法上的建议、《莱顿宣言》提出的关于合理使用定量评价的倡议、INORMS科技评价工作组提出的SCOPE方法，以及英国REF评估形成的影响力评价方法等。其中部分核心观点值得特别提及：① 科技评价是管理的需要，也是一把“双刃剑”，没有充足理由和充分准备不要轻易启动科技评价。② 科技评价离不开同行评议，难以通过简单量化方法评价。但是，量化方法如果使用得当，能够有效提升同行评议质量。③ 量化方法有好坏之分，在实际评价中需要辨别与选择好的量化方法，如学科规范化引文影响力（CNCI）指标[29]；避免使用不好的量化方法，如期刊影响因子。④ 需要关注支撑定量评价的数据质量及其来源的可靠性，没有数据质量及其来源可靠性保障，光是指标好看没有用。⑤ 通过使用唯一标识符（如ORCID），将科研产出、参与的科研活动等与科研人员绑定，在保证定量评价数据质量的基础上方便查寻获取，已经成为国际科学共同体越来越普遍的做法。

4.2 启示

以上通过中外比较得出的结论，对我国科技评价改革有许多启示，本文重点论述4个方面的启示。

（1）分类分步破“唯”。评价改革要明确责任主体和先后次序。目前我国的“四唯”中，“唯”要分类分步而破。① 严重性不同。“唯论文”“唯奖项”“唯帽子”比较严重需要重点关注，“唯学历”和“唯职称”也是问题但影响面较小。② 责任主体不同。“唯论文”的责任主体主要在科学共同体，这和国际科技评价改革一致；“唯奖项”“唯帽子”责任主体主要在政府，国际科技评价改革没有这方面的问题。这2个责任主体的问题是交互作用的，但是，有先后次序。“唯奖项”“唯帽子”本质是奖项和“帽子”过多，需要政府做减法，为科学共同体破“唯论文”、安心产出原创成果提供基础性管理制度保障。“唯论文”则需要借鉴国际科技评价改革经验，更好地发挥科学共同体自下而上改革的主动性，形成先行先试、敢为天下先的局面。

（2）慎重启动评价。评价是一把“双刃剑”，国际科技评价改革提倡不要轻易启动科技评价，这与我国科技评价改革节奏有相通之处。事实上，在破“唯”之前，我国首先开展的是“减量”改革，要求各单位清理过多过频、重复交叉的评价。但是，近年来减量风头过去后，评价冲动又有释放苗头；加上国家新出台的“全面实施预算绩效管理”的要求在落实中还没有和原有科技评价协调好，评价频次又多起来了。对此，需要通过建立对评价进行评价的机制约束评价冲动、规范评价制度和方法，而不仅仅停留于开展减量改革的阶段性运动。

（3）用好量化评价。作为同行评议的辅助方法，量化评价在本次国际科技评价改革中得到了重点关注，形成了较多共识。鉴于我国过去量化评价过于极端，在这次破“四唯”改革中，有一派观点主张完全放弃定量评价，回归同行评议。结合我国国情，这是不可取的。笔者曾经提出定量、定性相结合的BRIDGE理论[30]，主张通过表单化方法将数据材料和证据的隐性知识显性化，从而对同行评议起到支撑和约束作用。这一方面是将我国已有量化评价探索做到物尽其用，另一方面可能在定量定性结合的评价方法上形成改革突破，并为国际科技评价改革作出中国贡献。

（4）积极融入国际科技评价改革。目前，我国签署《旧金山宣言》的科研机构、高校和科学家个人还比较少，与我国庞大的科学共体还不相称。同时，我国还没有科研机构和高校加入国际《改革科研评价的协定》之中。这种情况与我国作为后发国家还存在一个逐步融入国际科学共同体的过程有关，也与我国科技评价改革由政府主导的特点有关，甚至近几年的疫

情也有较大影响。作为国际科学共同体的一部分，我国应该更加积极地融入国际科技评价改革。通过与国际科学共同体相互借鉴和促进，一方面可以更好地激发我国科学共同体在科技评价改革上的主动性，另一方面可以增加科学共同体之间的理解与信任，从而有利于加强全方位国际科技合作的纽带。

参考文献

1 徐芳, 晋新新, 吕雯洁, 等. 国际科研评价改革实践与启示——以DORA为代表的探索. 华南理工大学学报(社会科学版), 2022, 24(3): 94-100.
Xu F, Jin X X, Lyu W J, et al. International science and technology evaluation reform's practice and enlightenment—A exploration represented by DORA. Journal of South China University of Technology (Social Science Edition), 2022, 24 (3): 94-100. (in Chinese)

2 徐芳, 李晓轩. 科技评价改革十年评述. 中国科学院院刊, 2022, 37(5): 603-612.
Xu F, Li X X. Review on reform of research evaluation in past decade. Bulletin of Chinese Academy of Sciences, 2022, 37(5): 603-612. (in Chinese)

3 Garfield E. Citation indexes for science: A new dimension in documentation through association of ideas. Science, 1955, 122: 108-111.

4 盛小平, 毕畅畅, 唐筠杰. 国内外开放科学主题研究综述. 图书情报知识, 2022, 39(4): 101-113.
Sheng X P, Bi C C, Tang J J. A review of open science research at home and abroad. Documentation, Information & Knowledge, 2022, 39(4): 101-113. (in Chinese)

5 United Nations Educational, Scientific and Cultural Organization. UNESCO Recommendation on Open Science. (2021-11-09) [2023-12-26]. https://unesdoc. unesco. org/ark:/48223/pf000037994.

6 王飞跃, 缪青海. 人工智能驱动的科学研究新范式: 从AI4S到智能科学. 中国科学院院刊, 2023, 38(4): 536-540.
Wang F Y, Miao Q H. Novel paradigm for AI-driven scientific research: From AI4S to intelligent science. Bulletin of Chinese Academy of Sciences, 2023, 38(4): 536-540. (in Chinese)

7 Hicks D, Wouters P, Waltman L, et al. Bibliometrics: The Leiden Manifesto for research metrics. Nature, 2015, 520: 429-431.

8 Wilsdon J, Allen L, Belfiore E, et al. The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. (2015-07-06) [2023-12-26]. https://www. ukri. org/wp-content/uploads/2021/12/RE-151221-TheMetricTideFullReport2015.pdf.

9 The HuMetricsHSS Team. Humane Metrics Initiative. [2023-12-26]. https://humetricshss.org/humetricshss-principles/.

10 Wilsdon J, Bar-Ilan J, Frodeman R, et al. Next-generation metrics: responsible metrics and evaluation for open science. (2017-03-20) [2023-12-26]. https://www. researchgate. net/publication/315385216_Next-generation_metrics_responsi-ble_metrics_and_evaluation_for_open_science.

11 Moher D, Naudet F, Cristea I A, et al. Assessing scientists for hiring, promotion, and tenure. PLoS Biology, 2018, 16(3): e2004089.

12 INORMS-RWG. SCOPE Framework for Research Evaluation. [2024-01-04]. https://inorms. net/scope-framework-for-research-evaluation/.

13 European University Association. EUA Roadmap on Research Assessment in the Transition to Open Science. (2018-06-18) [2024-01-04]. https://eua. eu/downloads/publications/eua-roadmap-on-research-assessment-in-the-transition-to-open-science.pdf.

14 Dominik M, Badre A, Almas T A. et al. Publishing models, assessment, and open science. Report and outcomes from a workshop held by the Global Young Academy. (2018-10-29) [2023-12-26]. https://globalyoungacademy. net/wp-content/uploads/2018/10/APOS-Report-29.10.2018.pdf.

15 Global Research Council. Statement of Principles on Peer/Merit Review. [2023-12-26]. https://globalresearchcouncil. org/fileadmin//documents/GRC_Publications/French/GRC_2018_SoP_Declaration_de_Principes_sur_l_Analyse_Pair_Merite.pdf.

16 Moher D, Bouter L, Kleinert S, et al. The Hong Kong Principles for assessing researchers: Fostering research integrity. PLoS Biology, 2020, 18(7): e3000737.

17 Technopolis. Science Europe Study on Research Assessment Practices. (2020-02-04) [2023-12-26]. https://www.scienceeurope. org/our-resources/science-europe-study-on-research-assessment-practices/.

18 Global Research Council. Responsible Research Assessment —Global Research Council (GRC) Conference Report 2021. (2020-11-23)[2023-12-26]. https://globalresearchcouncil.org/fileadmin/documents/GRC_Publications/GRC_RRA_Conference_Summary_Report.pdf.

19 CommissionEuropean. Reforming research assessment: The Agreement is now final. (2022-07-20)[2023-12-26]. https://research-and-innovation. ec. europa. eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en.

20 Curry S, Rijcke S D, Hatch A, et al. The changing role of funders in responsible research assessment: Pprogress, obstacles and the way ahead. (2020-11-18) [2023-12-26]. https://rori.figshare.com/ndownloader/files/25518674.

21 Ghent University. Vision Statement for Evaluating Research at Ghent University. (2016-11-08)[2023-12-26]. https://www.ugent. be/en/research/research-strategy/evaluation/research-evaluation-principles.pdf.

22 Ghent University. Responsible use of quantitative indicators. (2023-10-9)[2024-01-04]. https://onderzoektips. ugent. be/en/tips/00001832/

23 Kozlov M. NIH plans grant-review overhaul to reduce bias. Nature, 2022, 612: 602-603.

24 National Institutes of Health. Biographical Sketch. [2023-12-26]. https://grants.nih.gov/grants/forms/biosketch.html.

25 Kozlov M. NIH issues a seismic mandate: Share data publicly. Nature, 2022, 602: 558-559.

26 徐芳, 刘文斌, 李晓轩. 英国REF科研影响力评价的方法及启示. 科学学与科学技术管理, 2014, 35(7): 9-15.

Xu F, Liu W B, Li X X. Research impact evaluation within UK REF and its implications. Science of Science and Management of S & T, 2014, 35(7): 9-15.(in Chinese)

27 Research and Innovation UK. Future Research Assessment Programme. (2022-12-09) [2023-12-26]. https://www. ukri. org/who-we-are/research-england/research-excellence/future-research-assessment-programme-frap/.

28 关于完善科技成果评价机制的指导意见. (2021-07-16) [2023-12-26]. https://www. gov. cn/zhengce/zhengceku/2021-08/02/content_5628987.htm?eqid=efd6cb8a0001575900000002646f098b.

29 郝若扬. 3种引文规范化指标RCR、CNCI和JNCI的相关性研究. 现代情报, 2023, 43(12): 133-142.

Hao R Y. Study on correlations among three standardized evaluation indices: RCR, CNCI and JNCI. Journal of Modern Information, 2023, 43(12): 133-142.(in Chinese)

30 徐芳, 李晓轩. 在“破”与“立”之间:科技评价的BRIDGE理论. 中国科学院院刊, 2022, 37(8): 1099-1107.

Xu F, Li X X. Between “Breaking” and “Building”: The BRIDGE Theory of research evaluation. Bulletin of Chinese Academy of Sciences, 2022, 37(8): 1099-1107. (in Chinese)

Review on reform of international research evaluation in past decade

XU Fang LI Xiaoxuan*

（1 Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;

2 School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, China;

3 S&T Management and Evaluation Committee of CASSSP, Beijing 100080, China）

Abstract In May 2013, with the release of the *San Francisco Declaration on Research Assessment* (DORA) as a symbol, the reform of international research evaluation began. Over the past decade, the reform of international research evaluation has undergone extensive theoretical preparation and initiatives, and is now moving towards practice. More than 350 institutions from over 40 countries have signed agreements to reform research evaluation and carried out pilot reforms, with initial results. This study systematically summarizes and analyzes the ten-year reform of international research evaluation in three aspects: reform goals, measures, and cases. From the perspective of comparing with China's research evaluation reform, research conclusions and inspirations are drawn.

Keywords international, research evaluation, reform, *San Francisco Declaration on Research Assessment* (DORA)

徐 芳 中国科学院管理创新与评估研究中心副主任，中国科学院科技战略咨询研究院创新研究员，科技管理国际联盟（INORMS）科技评价工作组委员。主要研究方向为科技管理与评价、科技政策学等，主要负责中国科学院研究所评估工作。
E-mail: xufang@casisd.cn

XU Fang Deputy Director of Management Innovation and Research Evaluation Center of Chinese Academy of Sciences (CAS), Associate Professor of Institutes of Science and Development, CAS, Member of the International Network of Research Management Societies. Dr. Xu's main research interests are S&T management and evaluation, S&T policy, etc., and she is mainly responsible for the evaluation of CAS institutes. E-mail: xufang@casisd.cn

李晓轩 中国科学院管理创新与评估研究中心主任，中国科学院科技战略咨询研究院研究员。《科研评价的旧金山宣言》（DORA）执行委员。主要研究领域为科研管理，涉及科技评价、科技人力资源管理、科研经费管理等研究方向。
E-mail: xiaoxuan@casisd.cn

LI Xiaoxuan Director of Management Innovation and Research Evaluation Center of Chinese Academy of Sciences (CAS), Professor of Institutes of Science and Development, CAS. Executive Member of the *Declaration on Research Assessment*. His research focuses on scientific research management, including research evaluation, human resource management, and research funding management. E-mail: xiaoxuan@casisd.cn

■责任编辑：岳凌生

*Corresponding author